颜国平，腾讯云－天御系统研发负责人。

1. 一直负责腾讯自有验证码、业务安全、防刷、账号安全等研发工作。
2. 内部支持的产品（游戏、电商、腾讯投资的O2O企业）非常广泛。
3. 在业务安全领域项目经验丰富，并且具备深度学习、大数据架构搭建等实战经验。

**一、背景介绍**

最近1~2年电商行业飞速发展，各种创业公司犹如雨后春笋大量涌现，商家通过各种活动形式的补贴来获取用户、培养用户的消费习惯。

但任何一件事情都具有两面性，高额的补贴、优惠同时了也催生了“羊毛党”。

“羊毛党”的行为距离欺诈只有一步之遥，他们的存在严重破环了活动的目的，侵占了活动的资源，使得正常的用户享受不到活动的直接好处。

今天主要分享下腾讯自己是如何通过大数据、用户画像、建模来防止被刷、恶意撞库的。

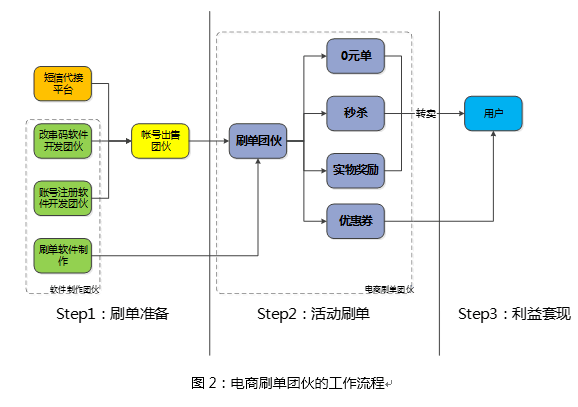
**二、黑产现状介绍**

“羊毛党”一般先利用自动机注册大量的目标网站的账号，当目标网站搞促销、优惠等活动的时候，利用这些账号参与活动刷取较多的优惠，最后通过淘宝等电商平台转卖获益。

一、羊毛党分工

他们内部有着明确的分工，形成了几大团伙，全国在20万人左右：

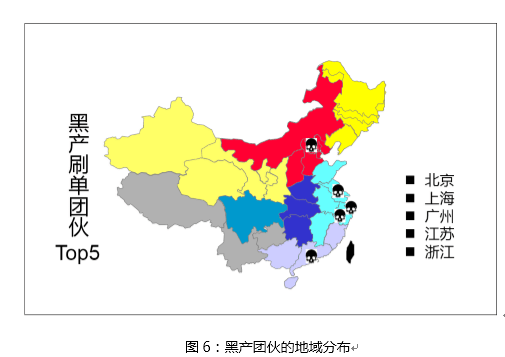
1. 软件制作团伙：专门制作各种自动、半自动的黑产工具，比如注册自动机、刷单自动机等；他们主要靠出售各种黑产工具、提供升级服务等形式来获利。
2. 短信代接平台：实现手机短信的自动收发，其实一些平台亦正亦邪，不但提供给正常的商家使用，一些黑产也会购买相关的服务。
3. 账号出售团伙：他们主要是大量注册各种账号，通过转卖账号来获利；该团伙与刷单团伙往往属于同一团伙。
4. 刷单团伙：到各种电商平台刷单，获取优惠，并且通过第三方的电商平台出售优惠，实现套现。



二、“羊毛党”从业特点

这些黑产团队，有三个特点：

1. 专业化：专业团队、人员、机器来做。
2. 团伙化：黑产已经形成一定规模的团伙，而且分工明确；从刷单软件制作、短信代收发平台、电商刷单到变卖套现等环节，已经形成完整的刷单团伙。
3. 地域化：黑产刷单团伙基本分布在沿海的一些经济发达城市，比如，北京、上海、广东等城市，这或许跟发达城市更加容易接触到新事物、新观念有关。



三、对抗刷单的思路

对抗刷单，一般来讲主要从三个环节入手：

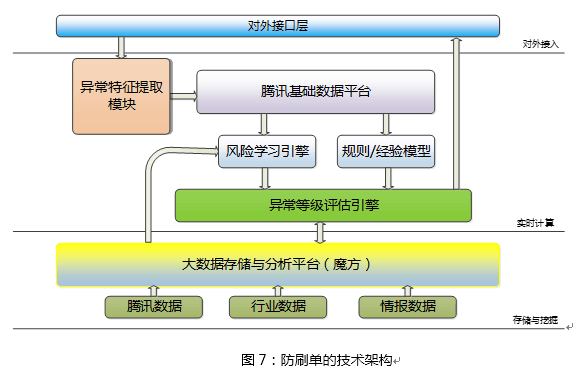
1. 注册环节：识别虚假注册、减少“羊毛党”能够使用的账号量。在注册环节识别虚假注册的账号，并进行拦截和打击。
2. 登录场景：提高虚假账号登录门槛，从而减少能够到达活动环节的虚假账号量。比如，登录环节通过验证码、短信验证码等手段来降低自动机的登录效率，从而达到减少虚假账号登录量、减轻活动现场安全压力的目的。
3. 活动环节：这个是防刷单对抗的主战场，也是减少“羊毛党”获利的直接战场；这里的对抗措施，一般有两个方面：

1）通过验证码（短信、语音）降低黑产刷单的效率。

2）大幅度降低异常账号的优惠力度。

**三、腾讯内部防刷架构**

一、腾讯内部防刷的架构图



二、模块详细介绍

1、风险学习引擎

风险学习引擎：效率问题。由于主要的工作都是线下进行，所以线上系统不存在学习的效率问题。线上采用的都是C++实现的DBScan等针对大数据的快速聚类算法，基本不用考虑性能问题。

风险学习引擎：采用了黑/白双分类器风险判定机制。之所以采用黑/白双分类器的原因就在于减少对正常用户的误伤。

例如，某个IP是恶意的IP，那么该IP上可能会有一些正常的用户，比如大网关IP。

再比如，黑产通过ADSL拨号上网，那么就会造成恶意与正常用户共用一个IP的情况。

黑分类器：根据特征、机器学习算法、规则/经验模型，来判断本次请求**异常**的概率。

白分类器：判断属于**正常**请求的概率。

2、矩阵式逻辑框架

我们以黑分类器为例来剖析下分类器的整个逻辑框架。

总的来讲我们采用了矩阵式的逻辑框架，最开始的黑分类器我们也是一把抓，随意的建立一个个针对黑产的检测规则、模型。

结果发现不是这个逻辑漏过了，而是那个逻辑误伤量大，要对那一类的账号加强安全打击力度，改动起来也非常麻烦。

因此我们就设计了这个一个矩阵式的框架来解决上述问题。

矩阵的横向采用了Adaboost方法，该方法是一种迭代算法，其核心思想是针对同一个训练集训练不同的弱分类器，然后把这些分类器集合起来，构成一个最终的分类器。

而我们这里每一个弱分类器都只能解决一种帐号类型的安全风险判断，集中起来才能解决所有账户的风险检测。

那么在工程实践上带来三个好处：

1. 便于实现轻重分离，比如某平台虚假账号集中在邮箱账号，策略就可以加大对邮箱账号的打击力度，影响范围也局限在邮箱帐号，而不是该平台所有的账号。
2. 减少模型训练的难度，模型训练最大的难度在于样本的均衡性问题，拆分成子问题，就不需要考虑不同账号类型之间的数据配比、均衡性问题，大大降低了模型训练时正负样本比率的问题。
3. 逻辑的健壮性，某一个分类器的训练出现了问题，受影响的范围不至于扩展到全局。

矩阵纵向采用了Bagging方法，该方法是一种用来提高学习算法准确度的方法，该方法在同一个训练集合上构造预测函数系列，然后以一定的方法将他们组合成一个预测函数，从而来提高预测结果的准确性。

**上面讲的部分东西，理解起来会比较艰涩，这里大家先理解框架，后续再理解实现细节。**

**四、腾讯大数据收集纬度**

大数据一直在安全对抗领域发挥着重要的作用，从我们的对抗经验来看，大数据不仅仅是数据规模很大，而且还包括两个方面：

1. 数据广度：要有丰富的数据类型。比如，不仅仅要有社交领域的数据、还要有游戏、支付、自媒体等领域的数据，这样就提供了一个广阔的视野让我们来看待黑产的行为特点。
2. 数据深度：黑产的对抗。我们一直强调纵深防御，我们不仅仅要有注册数据，还要有登录，以及账号的使用的数据，这样我们才能更好的识别恶意。

所以想要做风控和大数据的团队，一定要注意在自己的产品上多埋点，拿到足够多的数据，先沉淀下来。

**五、腾讯大数据处理平台－魔方**

我们的团队研发了一个叫魔方的大数据处理和分析的平台，底层我们集成了MySQL、MongoDB，Spark、Hadoop等技术，在用户层面我们只需要写一些简单的SQL语句、完成一些配置就可以实现例行分析。

这里我们收集了社交、电商、支付、游戏等场景的数据，针对这些数据我们建立一些模型，发现哪些是恶意的数据，并且将数据沉淀下来。

沉淀下来的对安全有意义的数据，一方面就存储在魔方平台上，供线下审计做模型使用；另一方面会做成实时的服务，提供给线上的系统查询使用。

一、腾讯用户画像沉淀方法

画像，本质上就是给账号、设备等打标签。

用户画像 ＝ 打标签

我们这里主要从安全的角度出发来打标签，比如IP画像，我们会标注IP是不是代理IP，这些对我们做策略是有帮助的。

以QQ的画像为例，比如，一个QQ只登录IM、不登录其他腾讯的业务、不聊天、频繁的加好友、被好友删除、QQ空间要么没开通、要么开通了QQ空间但是评论多但回复少，这种号码我们一般会标注QQ养号（色情、营销），类似的我们也会给QQ打上其他标签。

标签的类别和明细，需要做风控的人自己去设定，比如：地理位置，按省份标记。性别，安男女标记。其他细致规则以此规律自己去设定。

我们看看腾讯的IP画像，沉淀的逻辑如下图：

一般的业务都有针对IP的频率、次数限制的策略，那么黑产为了对抗，必然会大量采用代理IP来绕过限制。

既然代理IP的识别如此重要，那我们就以代理IP为例来谈下腾讯识别代理IP的过程。

识别一个IP是不是代理IP，技术不外乎就是如下四种：

1. 反向探测技术：扫描IP是不是开通了80,8080等代理服务器经常开通的端口，显然一个普通的用户IP不太可能开通如上的端口。
2. HTTP头部的X\_Forwarded\_For：开通了HTTP代理的IP可以通过此法来识别是不是代理IP；如果带有XFF信息，该IP是代理IP无疑。
3. Keep-alive报文：如果带有Proxy-Connection的Keep-alive报文，该IP毫无疑问是代理IP。
4. 查看IP上端口：如果一个IP有的端口大于10000，那么该IP大多也存在问题，普通的家庭IP开这么大的端口几乎是不可能的。

以上代理IP检测的方法几乎都是公开的，但是盲目去扫描全网的IP，被拦截不说，效率也是一个很大的问题。

因此，我们的除了利用网络爬虫爬取代理IP外，还利用如下办法来加快代理IP的收集：通过业务建模，收集恶意IP（黑产使用代理IP的可能性比较大）然后再通过协议扫描的方式来判断这些IP是不是代理IP。每天腾讯都能发现千万级别的恶意IP，其中大部分还是代理IP。

二、腾讯用户画像类别概览

三、防御逻辑

实时系统使用C/C++开发实现，所有的数据通过共享内存的方式进行存储，相比其他的系统，安全系统更有他自己特殊的情况，因此这里我们可以使用“有损”的思路来实现，大大降低了开发成本和难度。

数据一致性，多台机器，使用共享内存，如何保障数据一致性？

其实，安全策略不需要做到强数据一致性。

从安全本身的角度看，风险本身就是一个概率值，不确定，所以有一点数据不一致，不影响全局。

但是安全系统也有自己的特点，安全系统一般突发流量比较大，我们这里就需要设置各种应急开关，而且需要微信号、短信等方式方便快速切换，避免将影响扩散到后端系统。

四、接入系统

适应的场景包括：

* 电商o2o刷单、刷券、刷红包
* 防止虚假账号注册
* 防止用户名、密码被撞库
* 防止恶意登录

**Q&A**

**Q：风险学习引擎是自研的，还是使用的开源库？**

风险学习引擎包括两个部分，线上和线下两部分：

* 线上：自己利用c/c++来实现。
* 线下：涉及利用python开源库来做的，主要是一些通用算法的训练和调优。

**Q：请问魔方平台中用到的MongDB是不是经过改造？因为MongDB一直不被看好，出现问题也比较多。**

我们做了部分改造，主要是DB的引擎方面。

**Q：请问黑分类器和白分类器有什么区别？**

白分类器主要用来识别正常用户，黑分类器识别虚假用户。

**Q：风险概率的权重指标是如何考虑的？**

先通过正负样本进行训练，并且做参数显著性检查；然后，人工会抽查一些参数的权重，看看跟经验是否相符。

**Q：安全跟风控职责如何区分呢？**

相比安全，风控的外延更丰富，更注重宏观全局；针对一个公司来讲，风控是包括安全、法务、公关、媒体、客服等在内一整套应急处理预案。

**Q：如果识别错了，误伤了正常用户会造成什么后果么？比如影响单次操作还是会一直失败。**

如果识别错了正常用户不会被误伤，但是会导致体验多加了一个环节，如弹出验证码、或者人工客服核对等。